A PROPHET INEQUALITY
FOR INDEPENDENT RANDOM VARIABLES
WITH FINITE VARIANCES

by

D.P. Kennedy*

University of Cambridge

and

R.P. Kertz**

Georgia Institute of Technology

Abstract It is demonstrated that for each $n \geq 2$ there exists a universal constant, c_n, such that for any sequence of independent random variables $\{X_r,\ r \geq 1\}$ with finite variances, $\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_T \mathbb{E} X_T \leq c_n \sqrt{n-1} \max_{1 \leq i \leq n} \sqrt{\mathbb{V}\mathrm{ar}\,(X_i)}$, where the supremum is over all stopping times T, $1 \leq T \leq n$. Furthermore, $c_n \leq 1/2$ and $\liminf_{n \to \infty} c_n \geq 0.439485\ldots$.

1980 *Mathematics Subject Classification*: Primary 60G40.

Key words and phrases: Optimal stopping, prophet inequality.

* Statistical Laboratory, Univ. of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, UK.

** School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA.

Supported in part by NSF grant DMS 92-09586

1. Introduction

The terminology 'prophet' inequality is used to describe an inequality linking the quantity $\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right]$, the expected reward for a 'prophet' with foresight of the future, and $\sup_{T\in\mathcal{T}_n} \mathbb{E}X_T$, the optimal expected reward for a gambler who must use non-anticipating stopping times to stop at, or before time n, which holds for all sequences $\boldsymbol{X} = (X_1, X_2, \ldots)$ of integrable random variables in some class. In this optimal stopping reward for the gambler, the supremum is taken over the set $\mathcal{T}_n$ of stopping times T, $1 \leq T \leq n$, relative to the natural filtration, $\{\mathcal{F}_r,\ r \geq 1\}$, of the sequence $\boldsymbol{X}$. For example, for the class of all sequences of independent random variables, the 'ratio' prophet inequality of Krengel, Sucheston and Garling (cf. Krengel and Sucheston (1978)) states that

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] \leq 2 \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T, \tag{1}$$

for all sequences $\boldsymbol{X}$ of independent non-negative random variables; and the 'difference' prophet inequality of Hill and Kertz (1981) states that

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq \frac{1}{4}(b-a), \quad \text{for all} \quad n \geq 2, \tag{2}$$

for all sequences $\boldsymbol{X}$ of independent bounded random variables all taking values in some finite interval $[a,\, b]$. For these inequalities, the respective numbers '2' and '1/4' are universal constants for these respective collections of sequences $\boldsymbol{X}$, and the inequalities are sharp for each n ≥ 2. A survey of prophet inequalities may be found in Hill and Kertz (1992).

For prophet inequalities, it is of interest to find those (nontrivial) games for which the prophet realizes the greatest possible advantage over the gambler, that is, those (non-degenerate) sequences of random variables $\boldsymbol{X}$ within the collection for which equality for the prophet inequality is attained or nearly attained. These will generally be games in which certain of the random variables in the sequence $\boldsymbol{X}$ have as much variability as is allowed within the given collection, to be used by the prophet to gain advantage over the gambler. For example, equality is attained in (2) for the sequence $\boldsymbol{X}$ with $X_i \equiv \frac{1}{2}(b-a)$, $i = 1, \ldots, n-1$ and $X_n = b$ with probability $1/2$ and $X_n = a$ with probability $1/2$.

The purpose of this paper is to quantify the dependence between the prophet's advantage over the gambler and the variance of the random variables in the sequence $\boldsymbol{X}$, for a natural collection of sequences $\boldsymbol{X}$. Specifically, the issue considered is whether any information may be provided for the difference $\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T$ appearing in (2) when the requirement that the random variables are bounded is replaced by the assumption that the variances are finite. In this case it will be seen that the left-hand side of (2), which represents the advantage of the prophet over the gambler, may grow unboundedly with n and the object of this work is to determine the order of magnitude and to give a bound on that rate of growth. The principal result established is the following.

Theorem A *For each $n \geq 2$, there exists a universal constant, c_n, such that for any sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$ of independent random variables with $\mathbb{V}\mathrm{ar}\,(X_i) \leq \sigma^2 < \infty$, for each i, the inequality*

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq c_n \sigma \sqrt{n-1} \tag{3}$$

is satisfied; moreover, $c_n \leq 1/2$ and $\liminf_{n\to\infty} c_n \geq \sqrt{\ln(2) - 1/2} = 0.439485\ldots$.

This is a striking result! For the collection of sequences $\boldsymbol{X}$ in this theorem, the prophet can have a surprisingly large advantage. For 'ratio' and 'difference' prophet inequalities over different collections of sequences $\boldsymbol{X}$, of either nonnegative or bounded random variables, the prophet's advantage can be much larger over other collections of sequences, such as martingales, than over the collection of sequences of independent random variables. In contrast to inequalities (1) and (2), Dubins and Pitman (1980) and Hill and Kertz (1983) proved the ratio prophet inequality

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] \leq n \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T, \tag{4}$$

for all sequences $\boldsymbol{X}$ of nonnegative random variables, and the difference prophet inequality

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq \left(\frac{n-1}{n}\right)^n (b-a), \tag{5}$$

for all sequences $\boldsymbol{X}$ of random variables taking values in $[a,\, b]$. For these inequalities (4) and (5), the respective numbers 'n' and '$((n-1)/n)^n$' are again universal constants for these respective collections of sequences $\boldsymbol{X}$. The inequalities (4) and (5) are sharp for each n ≥ 2, and this sharpness can be demonstrated by using the subcollections of these collections consisting of martingale sequences. So for ratio and difference prophet inequalities, the prophet's advantage over the gambler can be large if sequences of martingales are considered, and relatively small for sequences of independent random variables. As Theorem A and the following Theorem indicate, this situation is reversed for these 'variance constrained' collections.

Theorem B *For each $n \geq 2$, there exists a universal constant, k_n, such that for any martingale sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$ with $\mathbb{V}\mathrm{ar}\,(X_i) \leq \sigma^2 < \infty$, for each i, the inequality*

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq k_n\sigma \tag{6}$$

is satisfied; moreover, $k_n \leq 1$ and the k_n's are increasing with $\lim_{n\to\infty} k_n = 1$.

Theorem B is a straightforward consequence of a martingale inequality given in Dubins and Schwartz (1988), and will be proved in Section 4.

2. Proof of Theorem A

The proof of Theorem A splits into two parts, the first of which establishes the existence of the required universal constant $c_n \leq 1/2$ which follows from the stronger inequality that

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq \frac{1}{2}\sqrt{\mathbb{V}\mathrm{ar}\,(X_2) + \cdots + \mathbb{V}\mathrm{ar}\,(X_n)} \tag{7}$$

for any independent random variables $X_1, X_2, \ldots, X_n$ and $n \geq 2$.

Verification of the inequality (7) is based on repeated application of two elementary inequalities – a conditional Jensen's inequality with the square-root function and inequality (10) below, an elementary inequality relating maxima and square roots. This application is carried out in a natural way within a supermartingale context.

To define this supermartingale, fix n and let v_r represent the optimal expected reward for stopping between times r and n, for $r = 1, \ldots, n$, so that

$$v_r = \sup\{\mathbb{E}X_T : T \in \mathcal{T}_n,\, T \geq r\}.$$

Note that $v_1 = \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T$, $v_n = \mathbb{E}X_n$, $v_r = \mathbb{E}\,(X_r \vee v_{r+1})$ for $r < n$, and $v_1 \geq \ldots \geq v_n$. Also let $M_r = \max_{1\leq i\leq r} X_i$ for $r = 1, \ldots, n$; let $\sigma_r^2 = \mathbb{V}\mathrm{ar}\, X_r$ for $r = 2, \ldots, n$ and impose the conventions that $v_{n+1} = -\infty = M_0$ and that an empty sum is zero.

Next define random variables $Z_1, \ldots, Z_n$ by setting

$$Z_r = v_2 \vee M_{r-1} + X_r \vee v_{r+1} + \sqrt{\sum_{i=r+1}^{n} \sigma_i^2 + \left(v_2 \vee M_{r-1} - X_r \vee v_{r+1}\right)^2}$$

for $r = 1, \ldots, n$; and observe that

(i) $Z_1, \ldots, Z_n$ is a supermartingale,

$$\text{(ii)}\quad Z_n = v_2 \vee M_{n-1} + X_n + \sqrt{\left(v_2 \vee M_{n-1} - X_n\right)^2} = 2\left(v_2 \vee M_n\right), \text{ and} \tag{8}$$

$$\text{(iii)}\quad Z_1 = v_2 + v_2 \vee X_1 + \sqrt{\sum_{i=2}^{n} \sigma_i^2 + \left(v_2 - v_2 \vee X_1\right)^2} \leq 2\left(v_2 \vee X_1\right) + \sqrt{\sum_{i=2}^{n} \sigma_i^2}.$$

Statements (8) (ii) and (8) (iii) are immediate, and (8) (i) follows since for $r = 2 \ldots, n$,

$$\begin{aligned}
&\mathbb{E}\big(Z_r \mid \mathcal{F}_{r-1}\big) \\
&\quad = v_2 \vee M_{r-1} + v_r + \mathbb{E}\left(\sqrt{\sum_{i=r+1}^{n} \sigma_i^2 + \left(v_2 \vee M_{r-1} - X_r \vee v_{r+1}\right)^2}\;\middle|\; \mathcal{F}_{r-1}\right) \\
&\quad \leq v_2 \vee M_{r-1} + v_r + \sqrt{\sum_{i=r+1}^{n} \sigma_i^2 + \mathbb{E}\left(\left(v_2 \vee M_{r-1} - X_r \vee v_{r+1}\right)^2 \;\middle|\; \mathcal{F}_{r-1}\right)} \\
&\quad = v_2 \vee M_{r-1} + v_r + \sqrt{\sum_{i=r+1}^{n} \sigma_i^2 + \mathbb{V}\mathrm{ar}\,\left(X_r \vee v_{r+1}\right) + \left(v_2 \vee M_{r-1} - v_r\right)^2} \qquad (9) \\
&\quad \leq v_2 \vee M_{r-1} + v_r + \sqrt{\sum_{i=r}^{n} \sigma_i^2 + \left(v_2 \vee M_{r-1} - v_r\right)^2} \\
&\quad = v_2 \vee M_{r-2} \vee X_{r-1} + v_r + \sqrt{\sum_{i=r}^{n} \sigma_i^2 + \left(v_2 \vee M_{r-2} \vee X_{r-1} - v_r\right)^2} \\
&\quad \leq Z_{r-1},
\end{aligned}$$

where the first inequality follows from the conditional Jensen's inequality; the second inequality from $\mathbb{V}\mathrm{ar}\,(X \vee a) \leq \mathbb{V}\mathrm{ar}\, X$ and the third inequality from the relation

$$a \vee b + c + \sqrt{d + (a \vee b - c)^2} \leq a + b \vee c + \sqrt{d + (a - b \vee c)^2} \tag{10}$$

which holds for real numbers a, b, c and d with $a \geq c$ and $d \geq 0$.

Now, inequality (7) is immediate, since

$$\mathbb{E}\left(M_n\right) \leq \frac{1}{2}\mathbb{E}Z_n \leq \frac{1}{2}\mathbb{E}Z_1 \leq v_1 + \frac{1}{2}\sqrt{\sum_{i=2}^{n}\sigma_i^2}$$

where the inequalities use (8) (ii), (8) (i) and (8) (iii) respectively.

The foregoing establishes that the quantity

$$c_n = \sup_{\boldsymbol{X}}\left[\frac{\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n}\mathbb{E}X_T}{\sqrt{n-1}\max_{1\leq i\leq n}\sqrt{\mathbb{V}\mathrm{ar}\left(X_i\right)}}\right] \leq \frac{1}{2},$$

where the supremum extends over all sequences $\boldsymbol{X}$ of independent random variables with finite variances.

The second part of the proof of the theorem involves showing the lower bound of $\liminf_{n\to\infty} c_n \geq \sqrt{\ln(2) - 1/2} = 0.439485\ldots$. To achieve this, produce appropriate examples through the following steps.

Step 1. For each $n \geq 2$, define as follows the family of finite sequences of independent random variables with X_1 being constant and $X_2, \ldots, X_n$ each taking two values; these random variables are dependent on n and on parameters $q_2, \ldots, q_n$ satisfying $0 < q_r < 1$ for $r = 2, \ldots, n$ and

$$\sqrt{\frac{q_{r+1}}{1-q_{r+1}}} \geq \frac{1}{\sqrt{q_r(1-q_r)}} \tag{11}$$

for $r = 2, \ldots, n-1$. Let $X_1 \equiv v_1$ and for $r = 2, \ldots, n$ set $X_r = u_r$ with probability $1 - q_r$ and $X_r = v_{r+1}$ with probability q_r, where the sequences $\{u_r\}$ and $\{v_r\}$ are required to satisfy $u_2 \leq u_3 \leq \ldots \leq u_n$, $v_1 = v_2 \geq v_3 \geq \ldots \geq v_n \geq v_{n+1} = 0$, and $u_r \geq v_{r+1}$ for $r = 2, \ldots, n$; further require that $v_r = \mathbb{E}\left(X_r \vee v_{r+1}\right) = \mathbb{E}X_r$ and that $\mathbb{V}\mathrm{ar}\left(X_r\right) = 1$ for $r = 2, \ldots, n$. These conditions imply that $\left(u_r - v_{r+1}\right)^2 q_r\left(1 - q_r\right) = 1$, giving

$$u_r = v_{r+1} + \frac{1}{\sqrt{q_r(1-q_r)}} \tag{12}$$

and

$$v_r = (1-q_r)u_r + q_r v_{r+1} = v_{r+1} + \sqrt{\frac{1-q_r}{q_r}}. \tag{13}$$

Thus, $\{u_r\}$ and $\{v_r\}$ are determined by the sequence $\{q_r\}$ by

$$u_r = \sum_{i=r+1}^{n} \sqrt{\frac{1-q_i}{q_i}} + \frac{1}{\sqrt{q_r(1-q_r)}} \quad \text{and} \quad v_r = \sum_{i=r}^{n} \sqrt{\frac{1-q_i}{q_i}}. \tag{14}$$

Note that, for $r = 2, \ldots, n-1$, $u_{r+1} \geq u_r$ is equivalent to inequality (11).

¿From the construction of these random variables, it follows that

$$\sup_{T \in \mathcal{T}_n} \mathbb{E} X_T = v_1 = v_2 = \sum_{i=2}^{n} \sqrt{\frac{1-q_i}{q_i}} \tag{15}$$

and

$$\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] = v_2 \prod_{r=2}^{n} q_r + \sum_{r=2}^{n} \left[u_r (1-q_r) \prod_{j=r+1}^{n} q_j \right], \tag{16}$$

with the convention that an empty product is 1.

Substitute for u_r and v_r from (12) - (14) to see that $\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_{T \in \mathcal{T}_n} \mathbb{E} X_T$ may be expressed as

$$\begin{aligned}
&\sum_{r=2}^{n} \left[\left(\prod_{j=r+1}^{n} q_j - \prod_{j=r}^{n} q_j \right) \left(\sqrt{\frac{q_r}{1-q_r}} + \sum_{i=r}^{n} \sqrt{\frac{1-q_i}{q_i}} \right) - \left(1 - \prod_{i=2}^{n} q_i \right) \sqrt{\frac{1-q_r}{q_r}} \right] \\
&\quad = \sum_{r=2}^{n} \left[\left(\prod_{j=r+1}^{n} q_j - \prod_{j=r}^{n} q_j \right) \sqrt{\frac{q_r}{1-q_r}} + \left(\prod_{j=r+1}^{n} q_j - 1 \right) \sqrt{\frac{1-q_r}{q_r}} \right] \\
&\quad = \sum_{r=2}^{n} \left[\left(\prod_{j=r}^{n} q_j + \prod_{j=r+1}^{n} q_j - 1 \right) \sqrt{\frac{1-q_r}{q_r}} \right].
\end{aligned}$$

It is now immediate that

$$c_n \geq \sup_{q_2, \ldots, q_n} \left\{ \frac{1}{\sqrt{n-1}} \sum_{r=2}^{n} \left[\left(\prod_{j=r}^{n} q_j + \prod_{j=r+1}^{n} q_j - 1 \right) \sqrt{\frac{1-q_r}{q_r}} \right] \right\}, \tag{17}$$

where the supremum in (17) extends over all $q_2, \ldots, q_n$, with $0 < q_i < 1$, which satisfy the inequality (11) for $r = 2, \ldots, n-1$.

Step 2. As follows, choose parameters $\{q_r\}$, depending on n, for the random variables of Step 1, and take limits to obtain the desired asymptotic bound. Let $g(\cdot)$ be a real-valued function on $[0,1]$ satisfying

$$g > 0, \quad g' \text{ is continuous and } (g'/g) + g^2 < 0 \text{ on } [0,1], \tag{18}$$

and for $r = 2, \dots, n$, set $q_r = 1 - (1/n)g^2((r-1)/n)$. In a straightforward way one checks that, for all n large, these numbers satisfy the conditions on the q_r's in Step 1; in particular, one uses the differential inequality of (18) to show that for all n large the inequalities of (11) hold for this choice of q_r's. Note that the differential inequality of (18) implies that g is strictly decreasing. For each such large n, define the sequence $X_1, \dots, X_n$ as in Step 1. For this choice of parameters, obtain from (15)-(17) that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{\sqrt{n}} &\left(\mathbb{E}\left[\max_{1\le i\le n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{r=2}^{n} g\left(\frac{r-1}{n}\right) \left(2\exp\left(-\frac{1}{n}\sum_{j=r}^{n} g^2\left(\frac{j-1}{n}\right)\right) - 1\right) \\
&= 2\int_0^1 g(u)\exp\left(-\int_u^1 g^2(s)ds\right)du - \int_0^1 g(u)du;
\end{aligned}$$

and hence that

$$\liminf_{n\to\infty} c_n \ge \sup_g \left[2\int_0^1 g(u)\exp\left(-\int_u^1 g^2(s)ds\right)du - \int_0^1 g(u)du \right], \tag{19}$$

where the supremum extends over all functions g satisfying (18).

Step 3. Calculus of variations techniques are used to show that the supremum on the right-hand side of (19) is no smaller than the number $\sqrt{\ln(2) - 1/2}$; this will complete the proof of the theorem. The approach used is directly analogous to that used in Section 5 of Assaf and Samuel-Cahn (1995) for a similar optimization problem, based on standard techniques found, for example, in Troutman (1983). To simplify the optimization and to set up the appropriate class of admissible functions, it is useful to reformulate the problem in terms of distribution functions $F(z)$ and (possibly degenerate) densities $f(z)$ through the identification

$$\begin{aligned}
F(z) &= \exp\left(-\int_z^1 g^2(s)ds\right) \quad \text{for } 0 \le z \le 1 \text{ and} \\
&\qquad F(z) = 0 \text{ for } z < 0 \text{ and } F(z) = 1 \text{ for } z > 1; \quad \text{and} \\
f(z) &= g^2(z)\exp\left(-\int_z^1 g^2(s)ds\right) \text{ for } 0 < z \le 1.
\end{aligned} \tag{20}$$

Thus, $F(z) = (1-p)F^c(z) + p\delta_0(z)$, where δ_0 is a point mass distribution function concentrated at zero, $p = \exp\left(-\int_0^1 g^2(s)ds\right)$ and $F^c(z)$ is a continuous distribution function concentrated on $(0,1)$ with density $(1-p)^{-1}f(z)$. With these identifications the optimization problem can be stated equivalently as follows:

$$\text{Maximize} \quad 2\int_0^1 \sqrt{f(z)F(z)}dz - \int_0^1 \sqrt{f(z)/F(z)}dz, \tag{21}$$

over positive functions $f(z)$ on $(0,1)$ satisfying

$$\int_0^1 f(z)dz \leq 1, \quad f'(z) \text{ is continuous, and } f(z)F(z) \text{ is strictly decreasing on } (0,1). \tag{22}$$

Define a functional $J(f)$ by $J(f) = \int_0^1 \left(2\sqrt{fF} - \sqrt{f/F}\right) dz$ for those positive functions f satisfying the first condition of (22), with associated F defined as in (20). One may show that for $\epsilon > 0$ and functions h with $f + \epsilon h$ in the domain of J,

$$\begin{aligned} L(f,h) &= \lim_{\epsilon \to 0} \epsilon^{-1} \left(J(f+\epsilon h) - J(f)\right) \\ &= \int_0^1 \left[\sqrt{\frac{f}{F}}\left(1 + \frac{1}{2F}\right) H + \sqrt{\frac{F}{f}}\left(1 - \frac{1}{2F}\right) h\right] dz, \end{aligned}$$

where H is defined through h just as F is defined through f. Thus, for an interior point f^* to be a local extremum of J, it is necessary that for all admissible functions h, $L(f^*,h) = 0$. It is possible to use as admissible functions h those integrable functions h for which $\int_0^1 h(z)dz = 0$. By considering functions $h_{u,\delta}(z)$, for $u \in (0,1)$ and $\delta > 0$ small, defined by $h_{u,\delta}(z) = I_{(u-\delta,u)}(z) - I_{(u,u+\delta)}(z)$, one obtains that for f to be an extremum of J,

$$0 = \lim_{\delta \to 0} \delta^{-2} L(f, h_{u,\delta}) = \sqrt{\frac{f}{F}}\left(1 + \frac{1}{2F}\right) - \frac{d}{du}\left[\sqrt{\frac{F}{f}}\left(1 - \frac{1}{2F}\right)\right],$$

and so,

$$\frac{1}{2}\sqrt{\frac{f}{F}}\left(1 + \frac{1}{2F}\right) + \frac{1}{2}\sqrt{\frac{F}{f}}\left(1 - \frac{1}{2F}\right)\frac{f'}{f} = 0. \tag{23}$$

Thus we seek f, and associated function F, satisfying (22) and (23). By rearranging and integrating equation (23), one obtains that F satisfies $F/(2F-1)^2 = k\,dF/du$ for some constant $k > 0$, and thus

$$u = k\left(2F^2(u) - 4F(u) + \ln F(u)\right) + C \tag{24}$$

for some constants $k > 0$ and C. The requirement that fF is strictly decreasing is seen to be equivalent to the condition that $F(z) > 1/2$ for $z \in (0,1)$. Now use (24), $F(1) = 1$, and the identification $p = F(0)$ to obtain $F^{-1}(y) = k(p)C(y) + 1$ for $y \in [p, 1]$ where $C(y) = 2(y-1)^2 + \ln y$ and $k(p) = -1/C(p)$. Although it is not possible to identify F, and the associated functions f and g, in closed form, one can calculate

$$\begin{aligned} J(f) &= \int_p^1 \sqrt{\frac{dF^{-1}}{dy}(y)} \left(2\sqrt{y} - \frac{1}{\sqrt{y}}\right) dy \\ &= \sqrt{k(p)} \int_p^1 \sqrt{4(y-1) + y^{-1}} \left(2\sqrt{y} - \frac{1}{\sqrt{y}}\right) dy = \sqrt{-C(p)}, \end{aligned}$$

and for $1/2 \leq p \leq 1$, this is maximized at $p = 1/2$ giving the value of $J(f) = \sqrt{\ln(2) - 1/2}$. The maximizer at $p = 1/2$ has the form $F^{-1}(y) = -\left(C(y)/C(1/2)\right) + 1$ for $1/2 \leq y \leq 1$.

3. Remarks on Theorem A

Several remarks are given concerning the statement and proof of Theorem A.

1. Note that the inequality

$$\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_{T \in \mathcal{T}_n} \mathbb{E} X_T \leq \frac{1}{2}\sigma\sqrt{n-1},$$

is sharp for $n = 2$ but one might not expect that it is sharp for $n \geq 3$. However, the inequality (7) from which it is deduced is sharp for all $n \geq 2$; take $X_i \equiv 1/2$, $i = 1, \ldots, n-1$ and $X_n = 1$ with probability $1/2$ and $X_n = 0$ with probability $1/2$. This inequality differs markedly from inequality (7) in the following fundamental way. In the ratio

$$\left(\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_{T \in \mathcal{T}_n} \mathbb{E} X_T\right) \Big/ \sqrt{\mathbb{V}\mathrm{ar}\,(X_2) + \cdots + \mathbb{V}\mathrm{ar}\,(X_n)},$$

the prophet can take advantage of the variances of $X_1, \ldots, X_{n-1}$ being small or zero and the variance of X_n being large; however, in the ratio

$$\left(\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_{T \in \mathcal{T}_n} \mathbb{E} X_T\right) \Big/ \sqrt{(n-1) \max_{2 \leq j \leq n} \mathbb{V}\mathrm{ar}\,(X_j)},$$

any variance penalizes the prophet by at least $(n-1)$ times that variance, so the prophet might as well take these $n-1$ variances to be the same, if this is possible. This is clear if one compares the example giving attainment for this inequality with the examples of the second part of Theorem A.

2. How did the choice of the supermartingale $\{Z_r\}$ in the first part of the proof of Theorem A arise? To obtain some insight into this choice, consider the first part of the inequality development in (9) for $r = n$. This is a conditional version of the following elementary inequality: for any random variable X with $\mathbb{E}X = \mu$ and $\mathbb{V}\mathrm{ar}\, X = \sigma^2$, and for any real number a,

$$\mathbb{E}\,(X \vee a) \leq \frac{1}{2}\left(a + \mu + \sqrt{\sigma^2 + (a-\mu)^2}\right), \tag{25}$$

(cf. Heijnen and Goovaerts (1989)). The formulation of the supermartingale $\{Z_r\}$ was based on initial use of the conditional version of (25) and the effect of repeated application of the same inequality and (10).

3. Sequences of two-valued random variables as constructed in Step 1 of the second part of the proof have been termed Bernoulli pyramids (cf. Hill and Kennedy (1992)). They occur frequently as extremal cases in prophet problems. Note that these sequences are supermartingales.

4. The lower bound in (17) may be derived explicitly for small values of n. For example, for $n = 3$, the maximizing values of q_2 and q_3 subject to the constraint (11) are $q_2 = 0.595373$ and $q_3 = 0.898017$ to give $c_3 \geq 0.466207$.

5. To gain insight into the convergence in Step 2 of the second part of the proof, one can consider the convergence of appropriate point processes. Define functions $V(r)$ and $U(r)$ on $[0,1]$ by $V(r) = \int_r^1 g(u)du$ and $U(r) = \int_r^1 g(u)du + (1/g(r))$, and the space $E = \{(r,x) : V(r) < x,\ 0 < r < 1\} \cup \{(1,0)\}$. Define point processes N_n, $n \geq 2$, and N on the space $M_p(E)$ of point measures on E by $N_n = \sum_{j=1}^n \epsilon_{(j/n, X_{n,j}/\sqrt{n})}$, where $X_{n,j} = X_j$, $j = 1, \ldots, n$, is as defined in Step 2 of the proof with $\epsilon_{(\cdot,\cdot)}$ denoting point mass at $(\cdot,\cdot)$, and $N = \xi + \epsilon_{(1,0)}$ where $\xi = \sum_i \epsilon_{(\tau_i, Y_i)}$ is the Poisson random measure on E with intensity measure λ given by $\lambda(A) = \int I\left((s, U(s)) \in A\right) g^2(s)ds$ for all Borel sets A in E. Observe that the differential inequality condition of (18) is equivalent to the condition that the function $U(r)$ is strictly increasing. Define the associated random variables M, T and

$Y(T)$ by $M = \max_i Y_i$ if $N \neq \epsilon_{(1,0)}$ and $M = V(0)$ if $N = \epsilon_{(1,0)}$; $T = \min_i \tau_i$ if $N \neq \epsilon_{(1,0)}$ and $T = 1$ if $N = \epsilon_{(1,0)}$; and $Y(T) = Y_i$ if $T = \tau_i$ and $Y(T) = 0$ if $T = 1$. Also define random variables T_n, $n \geq 2$, by $T_n = \min\{i < n : X_{n,i} = u_i\}$ if this set is non-empty and $T_n = n$ otherwise, and note that T_n is an optimal stopping time for $X_{n,1}, \ldots, X_{n,n}$. Then one can show the following convergence results hold by using standard techniques, found for example in Resnick (1987) and Serfozo (1982):

$$N_n \Rightarrow N \quad \text{in} \quad M_p(E);$$

$$\frac{1}{\sqrt{n}}\left(\max_{1\leq i\leq n} X_{n,i}\right) \Rightarrow M \quad \text{and} \quad \frac{1}{\sqrt{n}}\mathbb{E}\left(\max_{1\leq i\leq n} X_{n,i}\right) \to \mathbb{E}M;$$

$$T_n/n \Rightarrow T, \quad \frac{1}{\sqrt{n}}X_{T_n} \Rightarrow Y(T) \quad \text{and} \quad \frac{1}{\sqrt{n}}v_2 = \frac{1}{\sqrt{n}}\mathbb{E}\left(X_{T_n}\right) \to \mathbb{E}\left(Y(T)\right).$$

Here '$\Rightarrow$' denotes weak convergence in the appropriate space. To use the usual random measure contexts to model this convergence, we found it necessary to remove the boundary curve $\{(r, V(r)) : 0 \leq r < 1\}$ from the space E since this curve is attracting an infinite mass in the limit from the processes $\{N_n\}$ continuously along its length.

6. In the second part of the proof of Theorem A, the function $g(u)$ which gives the local extremum $J(f)$ value used there is not given in closed form. There are some natural choices of $g(u)$ which satisfy (18) or are at the 'boundary' of g's satisfying (18), which give values of $J(f)$ close to this value. For the function $g(u) = \sqrt{a/(b+u)}$, with $(a,b) = (0.17181, 0.0304828)$, the value of $J(f)$ is $0.43818\ldots$. For the variation in which $X_1 = \cdots = X_{r-1} = v_r$, and $X_r, \ldots, X_n$ are defined as in the proof with $g(u) = 0$ for $0 \leq u \leq \tau$ and $g(u) = \sqrt{\theta/u}$ for $\tau \leq u \leq 1$, where $r = r_n$ satisfies $r_n/n \to \tau$ as $n \to \infty$, one obtains for $\tau = (1/2)^{1/\theta}$ and $\theta = 0.146752\ldots$ that the value of $J(f)$ is $0.43485\ldots$.

7. We comment on some connections of this theorem to the literature.

(a) Growth of the difference $\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T$, as $n \to \infty$, was studied for i.i.d. random variables $X_1, \ldots, X_n$ in the domain of attraction of the extreme-value distributions in Kennedy and Kertz (1991), but prophet inequalities were not developed in this paper.

(b) For independent random variables $X_1, \ldots, X_n$ satisfying $X_n \geq 0$, $\mathbb{E}X_n = v_n$ and $\mathbb{V}\mathrm{ar}\,(X_n) = \sigma^2$, and for $j = 2, \ldots, n-1$, $X_j \geq v_{j+1}$, $\mathbb{E}X_j = v_j$ and $\mathbb{V}\mathrm{ar}\,(X_j) = \sigma^2$, for parameters $v_2 \geq \ldots \geq v_n$ of the theorem and $\sigma > 0$, one can use results of Stoyan (1973) concerning the convex order of probability measures to show that

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T \leq \mathbb{E}\left[\max_{1\leq i\leq n} X_i^*\right] - \sup_{T\in\mathcal{T}_n^*} \mathbb{E}X_T^*,$$

where $X_1^* = X_1$, the random variable X_n^* has mass at 0 of size $\sigma^2/\left(v_n^2+\sigma^2\right)$ and has continuous density on $((v_n^2+\sigma^2)/(2v_n), \infty)$ given by $f(x; v_n, \sigma) = \sigma^2/\left(2\left[(x-v_n)^2+\sigma^2\right]^{3/2}\right)$, and for $j = 2, \ldots, n-1$, the random variable X_j^* has mass at v_{j+1} of size $\sigma^2/((v_j - v_{j+1})^2 + \sigma^2)$ and has continuous density on $(v_{j+1} + \left\{(v_j - v_{j+1})^2 + \sigma^2\right\}/(2(v_j - v_{j+1})), \infty)$ given by $f(x;, v_j, \sigma) = \sigma^2/(2\left[(x-v_j)^2+\sigma^2\right]^{3/2})$. However, observe that for $j = 2, \ldots, n$, the random variable X_j^* has infinite variance. These X_j^*'s are connected to the sequence of random variables X_j constructed in the second part of the proof in that $\mathbb{P}\left(X_j^* = v_{j+1}\right) = \mathbb{P}\left(X_j = v_{j+1}\right)$ and $\mathbb{E}\left(X_j^* \mid X_j^* \neq v_{j+1}\right) = u_j$, and so X_j concentrates all the mass spread out by X_j^* within the interval $(v_{j+1} + \left\{(v_j - v_{j+1})^2 + \sigma^2\right\}/(2(v_j - v_{j+1})), \infty)$ on the one point $\mathbb{E}\left(X_j^* \mid X_j^* \neq v_{j+1}\right)$.

(c) Results on mean-variance type inequalities can be found in Section 6.4 of Tong (1990) and Pittenger (1990), and references therein.

4. Variance-Constrained Prophet Inequalities for Martingales

Proof of Theorem B: Dubins and Schwarz (1988) proved that the least upper bound, over all mean-zero martingales with variance bounded by v, of the expected essential supremum of the martingale is $\sqrt{v}$. The proof of Theorem B is an immediate consequence of this result. Let $\boldsymbol{X} = (X_1, X_2, \ldots)$ be any martingale sequence with $\mathbb{V}\mathrm{ar}\,(X_i) \leq \sigma^2 < \infty$ for each $i \geq 1$; then from the Optional Sampling Theorem (see e.g. Theorem 9.3.4 of Chung (1974)), $\sup_{T\in\mathcal{T}_n} \mathbb{E}X_T = \mathbb{E}X_1 = \ldots = \mathbb{E}X_n = \mu$, say, and so

$$\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T = \mathbb{E}\left[\max_{1\leq i\leq n} (X_i - \mu)\right] \leq \max_{1\leq i\leq n} \sqrt{\mathbb{V}\mathrm{ar}\,(X_i - \mu)} \leq \sigma.$$

This shows that the quantity

$$k_n = \sup_{\boldsymbol{X}} \left[\frac{\mathbb{E}\left[\max_{1\leq i\leq n} X_i\right] - \sup_{T\in\mathcal{T}_n} \mathbb{E}X_T}{\max_{1\leq i\leq n} \sqrt{\mathbb{V}\mathrm{ar}\,(X_i)}}\right] \leq 1,$$

where the supremum extends over all martingale sequences $\boldsymbol{X}$ of random variables with finite variances. One observes that $k_n \le k_{n+1}$, for example, by associating to each length-n martingale $X_1, \dots, X_n$ the length-$(n+1)$ martingale $Y_1 \equiv \mathbb{E}X_1, Y_2 = X_1, \dots, Y_{n+1} = X_n$.

To see that $\lim_{n\to\infty} k_n = 1$, consider the following martingales. For $n \ge 2$ fixed, the sequence $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ is such that the random variable X_j takes $(j+1)$ values for $j = 1, \dots, n$, and it is defined as follows: for $i = 1, \dots, n-1$ let $a_i = (n-i)\ln((n-i)/n) - (n-(i-1))\ln((n-(i-1))/n)$, and $b_i = -\ln((n-i)/n)$, then X_1 takes the values a_1 and b_1 with probabilities $1/n$, $(n-1)/n$ respectively, and for $j = 1, \dots, n-2$, if $X_j = b_j$, then

$$\mathbb{P}\left(X_{j+1} = a_{j+1} \mid X_1, \dots, X_j\right) = 1/(n-j) = 1 - \mathbb{P}\left(X_{j+1} = b_{j+1} \mid X_1, \dots, X_j\right);$$

for $i = 1, \dots, j$ if $X_j = a_i$, then $\mathbb{P}\left(X_{j+1} = a_i \mid X_1, \dots, X_j\right) = 1$; if $X_{n-1} = b_{n-1}$, then

$$\begin{aligned}\mathbb{P}\left(X_n = b_{n-1} + 1/p \mid X_1, \dots, X_{n-1}\right) &= p/e \\ &= 1 - \mathbb{P}\left(X_n = b_{n-1} - 1/(e-p) \mid X_1, \dots, X_{n-1}\right),\end{aligned}$$

and for $i = 1, \dots, n-1$, if $X_{n-1} = a_i$, as above, then $\mathbb{P}\left(X_n = a_i \mid X_1, \dots, X_{n-1}\right) = 1$, where p is the unique number in $(0,1)$ for which $\mathbb{E}\left(X_n^2\right) = 1$. Then $X_1, X_2, \dots$ is a martingale satisfying $\mathbb{E}X_j = 0$ and $\mathbb{V}\mathrm{ar}\,(X_n) = \mathbb{E}\left(X_n^2\right) = 1$. It follows that

$$\begin{aligned}k_n \ge \mathbb{E}\left[\max_{1\le i\le n} X_i\right] &= \left(\frac{n-1}{n}\right)\ln\left(\frac{n-1}{n}\right) + \frac{1}{n}\sum_{j=1}^{n-1}\left(-\ln\left(\frac{n-j}{n}\right)\right) + \frac{1}{ne} \\ &= \frac{1}{n}\ln\left(\frac{(n-1)^{n-1}}{(n-1)!}\right) + \frac{1}{ne};\end{aligned}$$

and one can use either Riemann-sum approximations or Stirling's formula to conclude that $\lim_{n\to\infty} k_n = 1$.

Finally, we give two remarks on Theorem B.

1. Dubins and Schwarz (1988) exhibit a stopped Brownian motion $(X_t)_{t\ge 0}$ for which

$$\mathbb{E}\left[\sup_{t\ge 0} X_t\right] - \sup_{T\in\mathcal{T}} \mathbb{E}X_T = \sqrt{\mathbb{V}\mathrm{ar}\,(X_\infty)}.$$

The continuous-time martingale of Dubins and Gilat (1978) also attains this equality. The martingale of the proof of this Theorem is a modification of the Dubins and Gilat martingale. Note that for the length-n martingale constructed in the proof of Theorem B, the variance of the terminal random variable is 1 and the variances of the previous random variables are strictly less than 1. This is in sharp contrast to the length-n process constructed in the second part of the proof of Theorem A, in which all but the first random variable have variance equal to 1.

2. For $n \geq 2$ fixed, let $\mathcal{X}_n$ denote the collection of martingale sequences $\boldsymbol{X} = (X_1, \ldots, X_n)$ satisfying $X_1 \equiv u_1$, and for $j = 1, \ldots, n-1$, if $X_j = u_j$ then

$$\mathbb{P}\left(X_{j+1} = u_{j+1} \,\middle|\, X_1, \ldots, X_j\right) = p_j = 1 - \mathbb{P}\left(X_{j+1} = 0 \,\middle|\, X_1, \ldots, X_j\right), \tag{26}$$

and if $X_j = 0$, then $\mathbb{P}\left(X_{j+1} = 0 \,\middle|\, X_1, \ldots, X_j\right) = 1$, for some $0 \leq u_1 \leq \ldots \leq u_n$, $p_j u_{j+1} = u_j$ for $j = 1, \ldots, n-1$, and $\mathbb{V}\mathrm{ar}\,(X_n) \leq 1$. For $[a, b] = [0, 1]$, equality is attained in (5) for the martingale sequence $\boldsymbol{X}$ in $\mathcal{X}_n$ of type (26) with $u_j = ((n-1)/n)^{n-j}$ for $j = 1, \ldots, n$ and $p_1 = \ldots = p_{n-1} = (n-1)/n$. However, one cannot hope to obtain a best-possible prophet advantage by using martingales from this class. For the sequence $\boldsymbol{X}$ in $\mathcal{X}_n$ of type (26), denote the arithmetic mean of $p_1, \ldots, p_{n-1}$ by $\overline{p}$, then one has

$$\begin{aligned}
\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] - \sup_{T \in \mathcal{T}_n} \mathbb{E} X_T &= u_1 \sum_{j=1}^{n-1} (1 - p_j) \\
&\leq \sqrt{\frac{\prod_{j=1}^{n-1} p_j}{1 - \prod_{j=1}^{n-1} p_j}} \sum_{j=1}^{n-1} (1 - p_j) = f(p_1, \ldots, p_{n-1}), \quad \text{say,} \\
&\leq \sqrt{\frac{(\overline{p})^{n-1}}{1 - (\overline{p})^{n-1}}} (n-1)(1 - \overline{p}) = f(\overline{p}, \ldots, \overline{p}) \\
&\leq \sup_{a > 0} \left[(n-1) a \left(1 - \left(\frac{a^2}{a^2 + 1}\right)^{1/(n-1)}\right)\right] \\
&\leq \sup_{a > 0} \left[a \ln\left(\frac{a^2 + 1}{a^2}\right)\right] = 0.804742\ldots < 1.
\end{aligned}$$

Here the first inequality uses the martingale property and $\mathbb{V}\mathrm{ar}\,(X_n) \leq 1$; the second inequality follows from the Schur-concavity of the function $f(p_1, \ldots, p_{n-1})$ on $[0, 1]^{n-1}$ (use Theorem A.4 of Chapter 3 of Marshall and Olkin (1979)); the third inequality uses the variable $a = \sqrt{p^{n-1}/(1 - p^{n-1})}$; and the fourth inequality uses the result that the function $g(n) = na\left(1 - (a^2/(a^2+1))^{1/n}\right)$ is increasing in n for $n \geq 1$ with $\lim_{n \to \infty} g(n) = a \ln\left((a^2+1)/a^2\right)$, for each $a > 0$.

References

[1] Assaf, D. and Samuel-Cahn, E. (1995). The secretary problem: minimizing the expected rank, with i.i.d random variables. *Adv. Appl. Prob.*, (to appear)

[2] Chung, K.L. (1974). *A Course in Probability Theory.* Second Edition. Academic Press, New York.

[3] Dubins, L.E. and Gilat, D. (1978). On the distribution of maxima of martingales. *Proc. Amer. Math. Soc.*, **68**, 337-338.

[4] Dubins, L.E. and Pitman, J. (1980). A maximal inequality for skew fields. *Z. Wahrsch. verw. Geb.*, **52**, 219-227.

[5] Dubins, L.E. and Schwarz, G. (1988). A sharp inequality for sub-martingales and stopping times. *Soc. Math. de France*, Astérisque, **157-158**, 129-145.

[6] Heijnen, B. and Goovaerts, M. J. (1989). Best upper bounds on risks altered by deductibles under incomplete information. *Scand. Actuarial. J.*, 23-46.

[7] Hill, T. P and Kennedy, D. P. (1992). Sharp inequalities for optimal stopping with rewards based on ranks. *Ann. Appl. Prob.*, **2**, 503-517.

[8] Hill, T. P. and Kertz, R. P. (1981). Additive comparisons of stop rule and supremum expectations of uniformly bounded independent random variables. *Proc. Am. Math. Soc.*, **83**, 582-585.

[9] Hill, T.P. and Kertz, R.P. (1983). Stop rule inequalities for uniformly bounded sequences of random variables. *Trans. Amer. Math. Soc.*, **278**, 197-207.

[10] Hill, T. P. and Kertz, R. P. (1992). A survey of prophet inequalities in optimal stopping theory. *Contemp. Math.*, **125**, 191-207.

[11] Kennedy, D.P. and Kertz, R.P. (1991). The asymptotic behavior of the reward sequence in the optimal stopping of i.i.d. random variables. *Ann. Prob.*, **19**, 329-341.

[12] Krengel, U. and Sucheston, L. (1978). On semiamarts, amarts, and processes with finite value. In *Probability on Banach Spaces*, Ed. by J. Kuelbs, Marcel Dekker, New York.

[13] Marshall, A.W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

[14] Pittenger, A.O. (1990). Sharp mean-variance bounds for Jensen-type inequalities. *Stat. and Prob. Letters*, **10**, 91-94.

[15] Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.

[16] Serfozo, R. (1982). Functional limit theorems for extreme values of arrays of independent random variables. *Ann. Prob.*, **10**, 172-177.

[17] Stoyan, D. (1973). Bounds for the extrema of the expected value of a convex function of independent random variables. *Studia Sci. Math. Hungar.*, **8**, 153-159.

[18] Tong, Y.L. (1990). *The Multivariate Normal Distribution.* Springer-Verlag, New York.

[19] Troutman, J.L. (1983). *Variational Calculus with Elementary Convexity.* Springer-Verlag, New York.